WHAT IS CLAIMED IS

1. Method for content-level monitoring, auditing, trending, and detection of anomalies in access to information, said information including electronic data on computers, said method comprising the steps of:
    a) Capturing of packets on the network
    b) Filtering packets to detect meaningful packets representing information content
    c) Decoding packets based on semantics of the application or protocol
    d) Analyzing packets to map message information contained in the packet into a quantitative representation
    e) Deriving a content signature from the quantitative representation
    f) Storing the content, along with the signature and attributes into a database
    g) Mining the content database to derive prototypical model of content, users, and time
    h) Detecting anomalies by finding strong deviations from the prototypical model
    i) Processing anomalies to minimize false alarms and increase the precision of anomalies
2. A method, according to claim 1, where the filtering is based on protocols and applications of interest
3. A method, according to claim 2, where protocols include database access protocol such as SQL, file server access protocol such as SMB, application protocols such as smtp, telnet, ftp, rcp, http, ldap, J2EE, .NET, etc. and applications include Notes, Documentum, Word, etc.

4. A method, according to claim 1, where the quantitative representation is captured as a content distribution vector that captures a frequency based distribution of key words in the message.
5. A method, according to claim 1, where the content signature is computed based on moment statistics such as the n-dimensional moment statistic
6. A method, according to claim 1, where the content signature is computed as a hash of the content
7. A method, according to claim 1, where the content signature is computed via document clustering where all documents that classify into one cluster share the same content signatures
8. A method, according to claim 1, where the attributes include user identity, location of access (source and destination IP address), time of access, content type (e.g. excel document vs. work document), content length, content hash, content encoding, content properties (e.g. ownership, time of creating, read/write/execute permissions, encryption, password protection status).
9. A method, according to claim 1, where mining may be based on statistical clustering and distance based metrics
10. A method, according to claim 9, where statistical metrics include frequency of all content signatures accessed by a user
11. A method, according to claim 9, where statistical metrics include time of all content signatures accessed by a user
12. A method, according to claim 9, where statistical metrics include location of all content signatures accessed by a user
13. A method, according to claim 1, where mining may be based on machine learning such as neural networks or rule-based expert systems
14. A method, according to claim 1, where mining may be augmented by content aging, where information is periodically deleted from the database

15. A method, according to claim 14, where aging depends on the nature of the mining algorithm, the organization, type of information being monitored, users, etc.
16. A method, according to claim 1, where anomalies are based on combinations of user, content, location, and time entities.
17. A method, according to claim 16, where anomaly is detected by a memory-based deviation where the content accessed by the user shows a deviation over the normal content accessed
18. A method, according to claim 16, where anomaly is detected by a rare content condition, where a user accesses content that is rarely accessed
19. A method, according to claim 16, where anomaly is detected by a time deviation where a user accesses content at a time different from historical access
20. A method, according to claim 16, where anomaly is detected by a location deviation where a user accesses content from a location different from historical access
21. A method, according to claim 1, where anomaly processing includes positive correlation with past security violation events
22. A method, according to claim 1, where anomaly processing includes negative correlation with past false alarms or non-events
23. A method, according to claim 1, where consistent anomalies are classified into pattern of misuse
24. A method, according to claim 1, where anomalies can be detected in real-time
25. A method, where anomaly detection is used for real-time protection of information
26. A method, according to 25, where real-time anomaly detection is used for protection via real-time alerts

27. A method, according to 25, where real-time anomaly detection is used for real-time protection via denial of access
28. A method, according to 25, where real-time anomaly detection is used for real-time protection via additional user validation
29. A method for correlating content, users, time, and space at the 'information' level, developing trends based on information access, and detecting anomalies of information access from confidential information repositories without requiring to know the specific type of information being accessed
30. A method, according to claim 29, where correlation is determined by identifying information that users consume, frequency with which information is accessed, time of access
31. A method, according to claim 29, where trends are used to identify prototypical or normal behavior of information access
32. A method, according to claim 29, where anomalies are identified by deviation from the normal behavior
33. A method, according to claim 29, where anomalies are identified by rare content events
34. A method, according to claim 33, where rare content events are used to identify critical information assets
35. A method, according to claim 29, where anomaly detection may be used for database retrievals, file server retrievals, application server retrievals, content scanning for email systems, or for anomaly detection for stored data content on end-user PCs and laptops
36. A method for content or information level anomaly detection that works when the content itself may be changing

37. A method, according to claim 36, which uses historical data and high level behavioral models such as memory and historical data to classify between anomalies and legitimate information access
38. A method for monitoring and auditing access to confidential information based on monitoring access behavior, characterizing access based on dimensions including user identity, location, time, and content, and detecting anomalies.
39. A method, according to claim 38, where content could be a table in a database, a file on a file server, a data object in an application server, a document in a document server, etc.
40. A method, according to claim 38, where auditing is used for privacy and legal compliance of regulations such as HIPAA, GLBA, CA 1386, etc. where an anomaly implies non compliance of these regulations.
41. An apparatus for monitoring, trending, and detection of anomalies in access to information, said critical information including electronic data on computers, comprises:

    a network based computing device that is used to capture packets, filter data content, decode packets based on protocol and application, derive content signatures, generate historical trends, detect anomalies, and provide real-time access control
42. An apparatus of claim 41, where it is implemented on a computing device and connected on a network as a passive tap
43. An apparatus claim 41, which is implemented as a network appliance that can derive information transparently without requiring logs
44. An apparatus of claim 41, where it is implemented on an end-user computing device such as a laptop of PC
45. An apparatus claim 41, where it is implemented as a shim on an application server

46. An apparatus of claim 41, where it is connected to systems monitoring consoles and user identity systems
47. An apparatus of claim 41, where it is connected to firewalls and other access control systems to enable real-time access control for anomalous information access
48. An apparatus claim 41, which is configured for compliance policies using a simple language